

Analyse de shifts dans des données industrielles de capteurs par AutoEncodeur Variationnel parcimonieux

Brendan L'Ollivier*, Sonia Tabti*, Julien Budynek*

*Fieldbox, Quai Armand Lalande, 33300, Bordeaux, France
blollivier@fieldbox.ai, stabti@fieldbox.ai, jbudynek@fieldbox.ai

Résumé. Cet article explore l'utilisation d'AutoEncodeurs Variationnels (VAE) parcimonieux dans le cadre de l'analyse des perturbations affectant la distribution de données industrielles, aussi appelées shifts. À cette fin, plusieurs modèles sont comparés, en particulier, nous introduisons le LassoVAE, un VAE avec décodeur parcimonieux dont la procédure d'entraînement est efficace en termes de temps de calcul. La comparaison s'appuie sur un protocole expérimental que nous avons mis en place qui inclut un générateur de données synthétiques simulant à la fois l'interaction parcimonieuse entre les variables d'un processus industriel et des shifts subis par ces dernières. Deux nouvelles métriques sont introduites afin d'évaluer chaque modèle sur sa capacité à isoler la source des shifts. Les résultats des expériences montrent la supériorité des modèles parcimonieux pour cette tâche.

1 Introduction

Le domaine du monitoring statistique de processus industriels regroupe l'ensemble des modèles statistiques et d'apprentissage automatique visant à surveiller l'état de fonctionnement d'un processus industriel à partir des données de capteurs (Joe Qin, 2003). La prise en compte de la nature multivariée des données industrielles en est l'un des principaux enjeux, et constitue toujours un sujet de recherche ouvert. Dans ce contexte, les modèles non-supervisés à facteurs latents ont été largement utilisés pour dé-corréler les variables de processus industriels (Qin et al., 2020). Ces modèles ont pour but de représenter les données observables par des combinaisons d'un nombre réduit (relativement à la dimension de l'espace observable) de facteurs latents indépendants. Chaque variable latente encode l'information contenue dans un groupe de variables corrélées, potentiellement interprétable en termes de systèmes physiques.

En particulier, l'analyse en composantes principales (PCA), qui extrait séquentiellement les facteurs qui expliquent le plus de variabilité dans les données, a reçu beaucoup d'attention pour la détection d'anomalies dans les processus industriels (Teppola et al., 1998; Yin et al., 2012; Qin et Chiang, 2019). L'efficacité de cette approche pour l'isolation de la source du shift repose sur l'interprétabilité des facteurs latents. Dans le cas d'interactions linéaires, l'interprétabilité est généralement fournie par les poids de la matrice de projection (Cadima et Jolliffe, 1995). Cependant, comme le signale (Greco et Farcomeni, 2016), la non-parcimonie de la matrice de poids peut conduire à une interprétation erronée des interactions entre facteurs